Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCES

JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000678

Generative Adversarial Networks for Downscaling Hourly Precipitation in the Canadian Prairies

Hebatallah M. Abdelmoaty[1,2], Simon Michael Papalexiou[1,3], Antonios Mamalakis[4,5], Shivam Singh[4,6], Vincenzo Coia[7,8], Melissa Hairabedian[7], Pascal Szeftel[7], and Patrick Grover[7]

[1]Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada, [2]Irrigation and Hydraulics Department, Faculty of Engineering, Cairo University, Giza, Egypt, [3]Institute for Global Water Security, Hamburg University of Technology, Hamburg, Germany, [4]Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA, [5]School of Data Science, University of Virginia, Charlottesville, VA, USA, [6]Environmental Institute, University of Virginia, Charlottesville, VA, USA, [7]BGC Engineering Inc., Victoria, BC, Canada, [8]Department of Civil and Environmental Engineering (DICA), Politecnico di Milano, Milan, Italy

Key Points:

- Wasserstein generative adversarial networks (WGAN) models maintained spatial and temporal dependencies, with the simplest structure performing best
- For Coupled Model Intercomparison Project phase 6 (CMIP6) downscaling, U-Network produced overly smoothed fields, while WGAN generated more localized and detailed patterns
- The WGAN downscaled precipitation accurately replicated the distribution of the Canadian Regional Climate Model (CanRCM4), effectively capturing extreme precipitation values

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. M. Abdelmoaty,
heba.abdelmoaty@ucalgary.ca

Citation:

Abdelmoaty, H. M., Papalexiou, S. M., Mamalakis, A., Singh, S., Coia, V., Hairabedian, M., et al. (2025). Generative adversarial networks for downscaling hourly precipitation in the Canadian Prairies. *Journal of Geophysical Research: Machine Learning and Computation*, *2*, e2025JH000678. https://doi.org/10.1029/2025JH000678

Received 6 MAR 2025
Accepted 7 OCT 2025

Abstract Developing robust downscaling methods is essential for maximizing the applicability of climate model outputs in engineering design and climate mitigation, particularly in a changing climate. This study evaluates four deep learning model configurations for downscaling, focusing on their structure, functionality, and ability to capture localized convective events in the Canadian prairies. These model configurations aim to downscale coarse-resolution climate model outputs (∼200 km) to the finer spatial resolution of regional climate models (∼50 km) for hourly precipitation. We introduce advanced metrics to assess the fidelity of precipitation downscaling, examining both marginal statistics and spatiotemporal dependencies. A U-Network (UNET) captures spatial and temporal dependencies efficiently while three generative adversarial networks (GANs) configurations incorporate a critic network to enhance the realism of generated fields. The study also evaluates the effects of a thresholding layer to constrain precipitation values and a convolution long short-term memory layer in the GAN critic to better capture temporal dependencies. Results show that all four model configurations effectively capture spatial dependencies, with the simplest GAN architecture outperforming others in preserving temporal dependencies. Latitudinal correlations are better preserved than longitudinal across all models. While UNET produces overly smoothed fields, GANs generate more detailed outputs when downscaling Coupled Model Intercomparison Project phase 6 projections. By optimizing deep learning models for this region, the study provides key insights into future precipitation trends, enabling the identification of localized storms. These findings are critical for improving infrastructure resilience across catchments in the prairies.

Plain Language Summary This study compares four deep learning models for downscaling future hourly precipitation data in the Canadian prairies. We focus on studying how well these deep learning models transform coarse climate projections (∼200 km) into finer, more detailed fields (∼50 km) for hourly precipitation fields. These models include a U-Network (UNET) and three generative adversarial network (GAN) configurations, assessed based on their ability to reproduce spatial patterns, temporal sequences, and statistical accuracy. While UNET captures these aspects satisfactorily, GANs produce more localized and detailed downscaled fields. The simplest GAN configuration preserves temporal patterns well. GANs also succeeded in downscaling Coupled Model Intercomparison Project phase 6 projections, confirming an increasing trend of hourly precipitation under the effect of climate change. The results show that deep learning can help downscale local storms and improve infrastructure planning.

1. Introduction

Global circulation models (GCMs) simulate the physical processes in the Earth's system, including water and energy cycles. The latest version of GCMs offered by the Coupled Model Intercomparison Project (CMIP6) shows a better representation of total precipitation than previous generations (Chen et al., 2020; Luo et al., 2021; Song et al., 2021). CMIP6 also offers a large ensemble of simulations through its scenario experiments, which examine potential climate outcomes based on varying levels of global warming and socioeconomic developments by the end of the century. These scenarios are categorized as shared socioeconomic pathways (SSPs), reflecting different global emission trajectories and their corresponding impacts (O'Neill et al., 2016). The commonly used SSPs include SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. Each SSP represents a different pathway: SSP1-2.6

© 2025 The Author(s). *Journal of Geophysical Research: Machine Learning and Computation* published by Wiley Periodicals LLC on behalf of American Geophysical Union.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

assumes a significant reduction in emissions, aligning with a sustainable development trajectory; SSP2-4.5 projects a continuation of historical emission trends; SSP3-7.0 represents a medium-to-high emission scenario associated with moderate socioeconomic challenges; and SSP5-8.5 envisions a future of rapid economic growth with high emissions, potentially leading to severe climate impacts. The availability of multiple scenarios and numerous realizations in CMIP6 provides a robust ensemble, which is extremely beneficial for assessing future climate risks and guiding mitigation strategies.

However, GCMs operate at coarse spatial resolutions (~200 km), which prevents them from accurately representing localized convective storms and the associated precipitation extremes (Dong et al., 2021). To obtain finer-scale precipitation fields, which are mostly the key drivers of extreme flood events, downscaling approaches are required. Regional climate models (RCMs) typically operate at 10–50 km resolution, while convection-permitting models (CPMs) can reach grid spacings of ≤5 km, allowing more explicit simulation of mesoscale convective processes (Chan et al., 2012; Schär et al., 2020). Although increasing resolution improves the representation of precipitation and reduces reliance on parameterized convection, it comes at substantial computational cost, limiting the spatial domains and ensemble sizes that can be simulated (Hwang, 2023; Li & Li, 2023). Consequently, RCMs are often employed as a compromise, offering improved spatial detail relative to GCMs while remaining computationally feasible for multidecadal simulations.

Leveraging the few RCM realizations, deep learning models (i.e., computational algorithms inspired by neural networks that can automatically detect complex patterns) can learn the spatial and temporal dependencies of fine-scale precipitation, enabling the numerical downscaling of the CMIP6 projection ensemble (Baño-Medina, 2020). Among the recent advances in deep learning models is the generative adversarial networks (GANs), which have proven competency in previous studies. Chaudhuri and Robertson (2020) investigated the ability of GAN models to downscale historical CMIP6 simulations between 1950 and 2010 for the Tsa Tue Bio-reserve, located in the Northwest Territories, Canada. The results illustrated that the downscaled daily precipitation aligns well with observations, including extreme precipitation. The same approach was used to downscale wind speed over the USA, highlighting a good performance for different configurations of GAN models. Additionally, GANs have the capacity to preserve temporal dependencies (Leinonen et al., 2021). Hess et al. (2023) presented a physically constrained GAN approach to correct biases in GCM outputs. Their method improved both local frequency distributions and spatial patterns of hourly precipitation over UK, addressing challenges in representing precipitation extremes.

GAN-based deep learning models are still being refined to achieve operational downscaling for GCMs (Sun et al., 2024). Li and Cao (2024) developed the MSG-SE-GAN model, which demonstrated an excellent performance in downscaling variables such as temperature and wind to a 6-hr temporal resolution over the western North Pacific region at 10 km. More recently, a variant of GANs, the Wasserstein GAN (WGAN) model, outperformed regression models for downscaling daily precipitation in New Zealand (Rampal et al., 2025). Iotti et al. (2025) introduced RainScaleGAN, a conditional deep convolutional GAN optimized for precipitation downscaling over European regions. Their model effectively outperformed the RainFARM model (Rebora et al., 2006), a widely used rainfall downscaling technique that applies a nonlinear transformation to a Gaussian random field, replicating spatial patterns and statistical distributions of daily precipitation using the ERA5-Land reanalysis data set.

Despite recent advancements in applying GAN-based models for downscaling hydroclimatic variables, there remains a research gap in applying these models to downscale CMIP6 projections to resolve fine-scale processes in hydrologically complex regions such as the Canadian prairies. Existing studies have demonstrated the potential of GANs in reproducing spatial and temporal dependencies, yet these efforts have typically focused on daily or coarser temporal resolutions, specific subregions, or reanalysis data sets rather than directly on CMIP6 GCM outputs. Notably, no prior studies have explored the use of GANs to downscale hourly precipitation from CMIP6 projections over the Canadian prairies. In this region, catchments are highly susceptible to extreme rainfall events and thus require high-resolution projections. Hence, informed risk assessment and adaptation planning could be implemented for protecting infrastructure crossing watersheds.

Accordingly, this study investigates the potential of four distinct deep learning models by incorporating several novel aspects: (a) the use of high temporal resolution hourly precipitation data, (b) the application of newly introduced advanced metrics to assess model performs in reproducing both marginal statistics and spatiotemporal dependencies across multiple spatial directions, and (c) the direct application of deep learning models to downscale CMIP6 projections in the Canadian prairies. These proposed models are tasked with downscaling GCM hourly

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 1. Study area map covering the study area covering the Canadian prairies.

precipitation outputs (~200 km) to the spatial resolution of RCMs (~50 km), in a case study focused on the Canadian prairies. We finally compare the downscaled data of the extreme hourly precipitation distribution derived from the coarse-resolution data with the high-resolution fields generated by the regional model. By achieving this purpose, hourly precipitation data can be available with ensembles under various climate change scenarios that are crucial for impact studies, risk management, and engineering designs, especially for vulnerable regional catchments.

2. Data

2.1. Study Area

In Canada, thunderstorms during the warmer seasons of spring and summer are associated with severe rainfall events typically lasting a few hours (Hanesiak et al., 2022; Milrad et al., 2011). These storms have led to sudden, localized flooding, overwhelming drainage systems, damaging critical infrastructure, and causing significant disruptions to agricultural productivity. Such events highlight the need for improved understanding of subdaily precipitation patterns. Additionally, with a changing climate, the frequency, intensity, and unpredictability of these extreme weather events across Canada are getting amplified, making them a growing concern for sustainable water management, urban planning, and disaster risk reduction (Brooks, 2013; Prein et al., 2017). This trend underscores the urgency of studying and preparing for these changes, particularly in regions already prone to such events. Given the susceptibility of this region to convective storms, we focus on the Canadian prairies as the study area to validate our downscaling schemes and explore precipitation projections under various SSPs, noting also their unique hydrological behaviour, driven by depressional storage and intermittent basin connectivity (Shook et al., 2021). The Canadian prairies region, characterized by its semiarid climate and reliance on agriculture, is particularly vulnerable to the effects of extreme precipitation (Mapfumo et al., 2023). The geographic extent of the study ranges from 121° to 94° west and from 48° to 61° north, considering a large portion of the Canadian prairies region (Figure 1).

2.2. Simulations

We use the most recent and publicly available model results of the Canadian RCM (CanRCM4) developed by Environment and Climate Change Canada (Diaconescu, 2016). CanRCM4 offers high-resolution simulations of hourly total precipitation at 0.44° (~50 km; Scinocca et al., 2016). For the historical period (1950–2005), CanRCM4 offers one historical simulation, which provides 20,416 days to train and test the four model

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

configurations. For prediction, we employ CNRM-ESM2-1 ensemble members from CMIP6 (∼200 km), including SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 projections spanning 2015–2100, to showcase the functionality of the employed methods. The selection of the downscaled spatial resolution (i.e., 50 km) was imposed by the availability of these simulations for the study area.

We use CanRCM4 outputs as the high-resolution reference data set within the study region. These fields consist of 1,404 grid cells (54 longitudes × 26 latitudes), corresponding to a resolution of approximately 50 km. To mimic GCM output resolution, we aggregate these fields to a coarser resolution of 98 grid cells (14 longitudes × 7 latitudes, ∼200 km) using the spatial average. Our deep learning models are trained to learn the mapping between low-resolution (200 km) and high-resolution (50 km) fields using CanRCM4 data. Once trained, the four model configurations are applied to downscale CMIP6 precipitation fields, which predominantly have the same coarse resolution used in this study (200 km). The input data consist of daily precipitation fields as a 3D structure (longitude × latitude × 24-hr time steps). The four deep learning models process these 3D inputs and generate higher-resolution (50 km) precipitation fields while preserving the temporal structure, effectively applying a super-resolution approach to downscaling.

3. Methods

3.1. Overview

A deep learning model is a multilayered neural network designed to automatically learn and extract hierarchical features from large and complex data sets to solve tasks like classification, prediction, or generation (Liu et al., 2017). A foundational component of many deep learning models is the convolutional neural network (CNN), which is well suited for capturing spatial features, patterns, and hierarchical structures within gridded data (Albawi et al., 2017; Bouvrie, 2006; O'Shea & Nash, 2015). To address problems involving both space and time, such as precipitation fields, convolutional long short-term memory (ConvLSTM) layers are often used. These layers combine the spatial capabilities of CNNs with memory mechanisms from recurrent networks to learn temporal dependencies, making them effective for modeling spatiotemporal processes.

Generative adversarial networks (GANs) are another class of deep learning models designed for data generation (Goodfellow et al., 2016). Conditional GANs, a variant where generation is conditioned on input data (e.g., low-resolution fields), are particularly suited for downscaling tasks (Mirza & Osindero, 2014). However, GANs often suffer from training instability and mode collapse, which can limit their effectiveness in generating diverse and realistic outputs (Wiatrak et al., 2020). To address these issues, the Wasserstein GAN (WGAN) introduces a different loss function based on the Wasserstein distance, providing smoother gradients and more stable training (Leinonen et al., 2021). WGANs consist of two competing neural networks: a generator and a critic. The generator is a network with a U-shaped architecture, known as a UNET model when applied as a stand-alone model. It also allows for incorporating a random noise input, which contributes to its stochastic nature. This generator attempts to produce high-resolution outputs from coarse-resolution inputs, while the critic evaluates the realism of the generated outputs by comparing them to the reference high-resolution data. This adversarial setup is particularly beneficial for downscaling precipitation, which involves complex, fine-scale spatial structures and high variability that traditional regression techniques struggle to reproduce (Rampal et al., 2024). WGANs are thus well suited for learning such intricate statistical distributions and producing physically plausible high-resolution precipitation fields (Annau et al., 2023; Chaudhuri & Robertson, 2020). An illustration of the conceptual design of WGANs is shown in Figure S1 in Supporting Information S1.

To investigate the effectiveness of various deep learning architectures for precipitation downscaling, four model configurations were developed and evaluated.

Model 1. The UNET model serves as a baseline for comparison. It employs an encoder-decoder structure with skip connections to capture spatial features of precipitation fields. The encoder reduces spatial dimensions while extracting hierarchical features, and the decoder reconstructs high-resolution outputs. This architecture is widely used for image-to-image translation tasks, and recently, it has dominated the field of weather forecasting (e.g., Kaparakis & Mehrkanoon, 2023; Li et al., 2024; Trebing et al., 2021).

Model 2. The second model (hereafter, WGAN) integrates a ConvLSTM layer into the generator of the WGAN. The generator maps low-resolution precipitation inputs to high-resolution outputs while capturing temporal dependencies. The critic evaluates the quality of generated fields without temporal layers that explicitly

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

understand the temporal dependence. However, the critic uses a set of fully connected layers that decompose and process the third dimension (i.e., time). The outputs of this model are not constrained to nonzero values. Note that we perform a thresholding in a postprediction setting, simply by setting all predicted values <0.1 to zero (See Section 3.4 for more details).

Model 3. Intermittent precipitation, where many grid cells contain near-zero or zero precipitation values, poses a challenge in downscaling because models may struggle to distinguish between wet and dry regions. This issue can lead to unrealistic precipitation patterns, particularly by introducing spurious low-intensity precipitation or failing to preserve the spatial distribution of dry areas. To mitigate this challenge, the third model (hereafter, WGAN_NZ) includes a nonzero thresholding layer within the generator in addition to the structure of the previous model. This layer ensures that generated precipitation fields do not include negative values or positive ones very close to zero, effectively distinguishing between zero and nonzero precipitation regions. The critic architecture remains the same as for Model 2.

Model 4. The fourth model (hereafter, WGAN_NZ_DT) combines the advantages of temporal and threshold-aware mechanisms. This model has the same description as Model 3 but with an additional ConvLSTM layer in the critic network. The nonzero thresholding layer remains in the generator.

3.2. Models Structure

The four models employed are permutations and combinations of the architecture shown in Figure 1. Central to this architecture are residual blocks, each consisting of two convolutional layers (ConvL) followed by activation layers. These blocks process the input through the activation and convolutional layers and then add the input back to the output. This design deepens the model without altering its overall structure. Inactive residual blocks with near-zero weights function as identity maps, passing their input forward without modification. During training, the network dynamically activates additional blocks as it learns to leverage deeper features. While increasing the number of residual blocks elevates computational cost, it does not compromise performance (Leinonen et al., 2021). The ConvLs in the residual blocks use a kernel size of 1×1, while ConvLs outside the residual blocks have a kernel size of 3×3.

The generator network (Figure 2a) is designed to produce high-resolution outputs from low-resolution inputs. To enhance realism and support ensemble generation, a noise factor is introduced and concatenated (i.e., merged) with the low-resolution input. The generator follows a UNET structure, serving as the first model in this study. In the encoding phase, the generator takes two inputs: a low-resolution sequence of 24 time steps and a latent noise vector sampled from a standard Gaussian distribution. The noise vector is transformed via a dense layer and broadcast across time to introduce stochasticity (i.e., a shared noise across the 24 time steps for each daily input). During training and inference, new noise vectors are sampled independently for each input, enabling the model to generate a range of plausible high-resolution outputs conditioned on the same low-resolution input. Moreover, by applying the same noise vector across all time steps for each daily input, the model incorporates global stochasticity at the daily scale without disrupting the temporal dependencies it is designed to capture. The decoding phase begins with a ConvLSTM layer to learn temporal dependencies, followed by a series of upsampling layers and ConvLs, ultimately producing the high-resolution output. To preserve edge continuity and prevent artifacts, reflection padding layers are applied after upsampling layers. The activation function used throughout the network is the leaky rectified linear unit (leaky ReLU) with a negative slope of 0.2, except for the output layer, which uses a linear activation function (Figure 2a). We chose a negative slope of 0.2 for the leaky ReLU activation based on the common practices, where this slope has been shown to provide stable gradients and improved convergence during adversarial training (e.g., Gulrajani et al., 2017; Leinonen et al., 2021; Radford et al., 2016).

An optional threshold layer is included after the output layer to prevent the generation of negative or very small positive nonzero values. This threshold layer is utilized in the WGAN_NZ and WGAN_NZ_DT models. An ablation study (i.e., a process used to determine the optimum configuration by changing parameters or adding/removing components in a network) revealed that dropout layers, typically used to prevent overfitting, were unnecessary for this model. Instead, a regularization technique with a decay weight of 0.001 was incorporated. The critic network (Figure 2b) starts with two independent pathways for the low-resolution and high-resolution inputs. Each input passes through several residual blocks before being concatenated and processed through an additional residual block. This merged output allows the critic to effectively leverage information from both inputs.

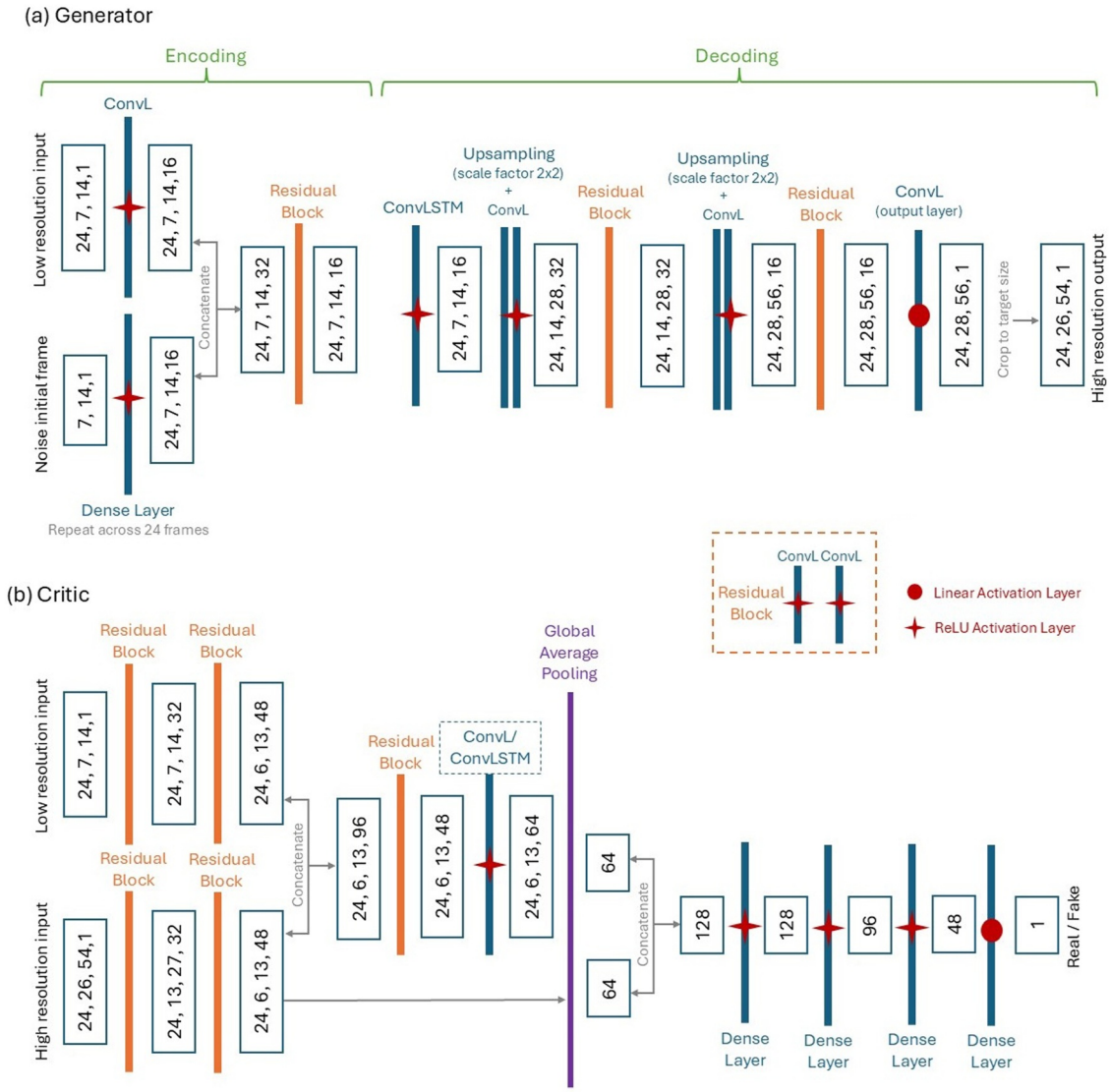

Figure 2. The structure of the generator and critic networks used for the U-Network and Wasserstein generative adversarial networks models. Each box represents the dimensions of the inputs as they pass through the layers. The dimensions are expressed as time steps, number of latitudes, number of longitudes, and number of channels. The single numbers indicate the channel numbers after flattening the other dimensions.

Next, the combined features pass through either a ConvL layer (used in WGAN and WGAN_NZ) to deepen the network's learning capacity or a ConvLSTM layer (used in WGAN_NZ_DT) to enhance temporal understanding. The combined fields are concatenated again with the high-resolution fields from earlier residual blocks, further improving the critic's ability to evaluate whether the generated high-resolution field is real or false. Afterward, the fields are flattened using a global average pooling layer and pass through several dense layers, producing a single output that determines whether the input field is real or fake. Similar to the generator, the critic employs the leaky ReLU activation function with a negative slope of 0.2, except for the output layer, which uses a linear activation function. No dropout layers are included; instead, spectral normalization is applied to all ConvLs. The iterative design process refined key architectural choices, including the number of residual blocks, convolutional layers, activation functions, regularization weights, and the inclusion of reflection padding and spectral normalization.

3.3. Training Procedure

The full data set of CanRCM4 comprises daily precipitation fields from 1950 to 2005, totaling 20,416 days. We used the first 13,611 days (approximately 1950–1987), representing 70% of the data, to train all deep learning

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

models, while the remaining 6,805 days (1987–2005) were reserved for evaluation. The split was performed chronologically to simulate a realistic scenario in which models are trained on historical climate data and evaluated on subsequent periods.

The UNET model is optimized by minimizing the mean absolute error (MAE), a straightforward loss function that directly compares predicted and true values. WGAN models are optimized using the Wasserstein loss, which approximates the distance between real and generated data distributions. UNET consists of a single network that directly learns to generate high-resolution outputs, whereas WGANs involve two interacting networks that learn adversarially. As a result, UNET models often converge more easily and require less tuning than WGANs. An overview of the training parameters for the UNET and WGAN models is provided in Table S1 in Supporting Information S1. The trained UNET model was then used to initialize the generator in the WGAN framework, which considerably enhanced training stability (Frégier & Gouray, 2021; Ham et al., 2020). Additionally, Figure S2a in Supporting Information S1 shows the UNET's training and validation loss, confirming its effective convergence using the MAE objective. Figure S2b in Supporting Information S1 presents the WGAN's adversarial training losses. Notably, while the critic loss remains relatively stable, the generator loss exhibits fluctuations that reflect adversarial learning. These dynamics, absent in the UNET-only training, demonstrate how the WGAN framework further refines the generator's outputs beyond MAE minimization, supporting the added value of adversarial training in capturing more realistic spatial structures.

For WGAN models, the training process involved optimizing the generator and critic losses using the Wasserstein loss with the gradient penalty (Gulrajani et al., 2017). It is an enhancement of the standard Wasserstein loss used in GANs. This loss stabilizes training and improves convergence by addressing the limitations of weight clipping, a technique used in the original Wasserstein GAN to constrain the critic's weights but often leading to poor optimization and capacity issues. The gradient penalty enforces the Lipschitz constraint, which ensures that the critic has gradients with a norm close to 1.

The Wasserstein objective function for the critic is defined as follows:

$$L_C = E_{z\sim P_g}\left[C(G(z))\right] - E_{x\sim P_r}\left[C(x)\right] \tag{1}$$

where $C(x)$ is the critic; $x \sim P_r$ represents real samples from the true data distribution; $z \sim P_g$ represents low-resolution samples merged with the input noise given to the generator; $G(z)$ is the generated high-resolution samples; and E is the expectation of the outputs. $C(x)$ is a continuous function that results in values $\in$ R, which represent a score measuring how real x is. Higher values indicate that a sample is more realistic, while lower values indicate that a sample is more fake.

The gradient penalty term is introduced as follows:

$$L_{GP} = \lambda \cdot E_{\hat{x}\sim P_{\hat{x}}}\left[\left(\|\nabla_{\hat{x}} C(\hat{x})\|_2 - 1\right)^2\right] \tag{2}$$

where $\hat{x}$ is a random interpolation between real and generated samples, defined as $\hat{x} = \varepsilon x + (1-\varepsilon)G(z)$, where $\varepsilon \sim$ Uniform(0, 1); $\|\nabla_{\hat{x}} C(\hat{x})\|_2$ is the L_2-norm of the gradient C with respect to $\hat{x}$; and λ is a weight coefficient that controls the weight of the gradient penalty.

The final objective function for the critic, including the gradient penalty, is as follows:

$$L_C = E_{z\sim P_g}\left[C(G(z))\right] - E_{x\sim P_r}\left[C(x)\right] + L_{GP} \tag{3}$$

For the generator, the loss is the negative expected value of the critic output on generated samples as follows:

$$L_G = -E_{z\sim P_g}\left[C(G(z))\right] \tag{4}$$

The generator aims to minimize the loss function by increasing the critic's score given to its generated samples. Since the critic network tries to maximize the difference between the score given to the real and generated samples while the generator aims to minimize that separation, this adversarial relationship is intentional. It is designed to drive the iterative improvement of both networks.

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The training parameters were determined through an iterative design process and common practice in GAN literature (e.g., Bhatia & Dahyot, 2019; Gulrajani et al., 2017; Leinonen et al., 2021; Radford et al., 2016; Tirel et al., 2024). The gradient penalty weight (λ) was optimized to a value of 10. The Adam optimizer was selected for its combined benefits of momentum-based optimization and adaptive learning rate adjustment. A learning rate of 0.0001 was chosen for the critic and 0.0002 for the generator, with $\beta_1 = 0.5$ and $\beta_2 = 0.9$ as the exponential decay rates for the first and second moment estimates in the optimizer function. The batch size was set to 32, and the model was trained over 30 epochs. To enhance performance, the critic was trained for more steps than the generator, as recommended in prior work (Leinonen et al., 2021). Specifically, in our model, the critic underwent one additional training step ahead of the generator during each iteration.

3.4. Evaluation Analyses

3.4.1. Distribution Evaluation

For each 24-hr time series, the probability of zero precipitation (P_0), mean, second L-moment, L-skewness, and L-kurtosis were calculated for both the simulated and original fields. Unlike many previous downscaling studies that rely primarily on mean and variance, our analysis incorporates higher-order L-moments (L-skewness and L-kurtosis), which provide a more rigorous assessment of the distributional shape and tails of hourly precipitation. We employed L-moments in our analysis because they provide a more robust evaluation of model performance than ordinary moments in reproducing the statistical characteristics of hourly precipitation. For instance, a bias in the mean indicates the model struggles with capturing the central tendency, while a bias in the second L-moment reveals issues with reproducing variability. Similarly, discrepancies in L-skewness or L-kurtosis suggest the model is misrepresenting the asymmetry or the peakedness and tails of the distribution, respectively. These statistics were assessed using scatterplots, along with bias (Equation 5) and root mean square error (RMSE; Equation 6) metrics, to quantify the differences between the simulated and original fields. We applied a precipitation threshold of 0.1 mm hr^{-1} to hourly data to exclude drizzle-like signals, those below the detection limits of standard rain gauges and model outputs, and to focus our statistical analysis on meaningful precipitation events. This threshold is commonly used in both observational and modeling studies to define "wet hours" and reduce noise from spurious light precipitation (e.g., Lauer et al., 2023; Lei et al., 2022; Thomassen et al., 2022). Accordingly, P_0 was calculated using the zero precipitation (i.e., values below 0.1 mm), while the remaining statistics were estimated using nonzero precipitation (i.e., values above 0.1 mm).

$$\text{Bias} = \frac{1}{n}\sum_{t=1}^{n}(S_s - S_o)_t \tag{5}$$

$$\text{RMSE} = \left(\frac{1}{n}\sum_{t=1}^{n}(S_s - S_o)_t^2\right)^{1/2} \tag{6}$$

where, S_s is the simulated statistic for a time series, S_o is the original statistic for the corresponding time series, and n is the number of time series across the study region. The study region, as represented in the high-resolution fields, consists of 1,404 grid cells, each containing a separate time series.

3.4.2. Spatial Similarity

The Pearson linear correlation coefficient (r; Equation 7) was calculated between a field and its shifted field in both the longitudinal (up to 10 lags, ~500 km) and latitudinal (up to 5 lags, ~250 km) directions. A shifted field refers to a spatially displaced version of the original field, offset by a specified lag in a given direction. This concept is analogous to temporal autocorrelation, which relates a value at a specific time to the value at the same location but at a different (shifted) time. Bias and RMSE metrics were then used to quantify the model performance for the simulated and original fields. In addition, we adopt the similarity index (SI; Equation 8), originally developed in image analysis, which has rarely been applied in precipitation downscaling studies, thereby extending the evaluation tool kit to capture structural fidelity of generated fields (Zhou et al., 2004). This index quantifies the similarity between the generated and original fields, with a value of 1 indicating perfect similarity and 0 indicating no similarity.

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \sum_{i=1}^{N}(O_i - \overline{O})^2}} \tag{7}$$

$$\mathrm{SI} = \frac{2 \times \sum_{i=1}^{N} X_i O_i}{\sum_{i=1}^{N} X_i^2 + \sum_{i=1}^{N} O_i^2} \tag{8}$$

where, N is the number of fields; X_i is the simulated field value; O_i is the original field value; $\overline{X}$ is the average of the simulated field value; and $\overline{O}$ is the average of the original field value.

3.4.3. Temporal Pattern

For each 24-hr time series, the Pearson linear autocorrelation coefficient (Equation 7) was calculated for lags up to 5 hr, where N is the number of individual time series within each field. Bias and RMSE metrics were then used to quantify the differences between the simulated and original fields. Additionally, to account for regional variability in individual grid time series, the average regional hourly precipitation from the original and generated simulations was compared using several metrics.

First, the coherence coefficient (CC; Equation 9) was calculated to measure the degree of linear correlation between two signals (simulated and original) at each frequency based on spectral density. To our knowledge, applying spectral coherence and nonparametric KS tests in the context of evaluating GAN-generated precipitation fields is a novel application, offering sensitivity to both frequency domain similarity and distributional shifts beyond mean/variance metrics.

$$\mathrm{CC} = \frac{|D_{XO}(f)|^2}{D_{XX}(f) \times D_{OO}(f)} \tag{9}$$

where, $D_{XO}(f)$ is the cross-spectral density between the simulated signal, X_i, and the original one, O_i; $D_{XX}(f)$ is the power spectral density of the simulated time series; and $D_{OO}(f)$ is the power spectral density of the original time series.

Second, the Kolmogorov-Smirnov test statistic (KS; Equation 10) was employed to compare the nonparametric cumulative distributions of the average daily time series from the simulated and original fields, where KS ranges from 0 to 1, with 0 indicating perfect match between the two distributions, and higher values indicate greater difference between the distributions.

$$\mathrm{KS} = \sup_x |F_X(x) - F_O(x)| \tag{10}$$

where, $F_X(x)$ is the simulated distribution; $F_O(x)$ is the original distribution; and $\sup_x$ is the supremum (maximum) over all values of x.

Finally, bias and RMSE were computed for the spatial average of the nonparametric autocorrelation functions (ACF) of the original and generated fields for each model.

3.4.4. Event-Based Evaluation

We employed event-based skill scores to reduce the influence of bias and assess whether specific events occurred at the same spatial and temporal locations in the predictions as in the original fields. First, both the original and generated fields were transformed into binary representations based on predefined thresholds. Specifically, we used a wet/dry threshold of 0.1 mm, the median, and the 75th percentile of the original fields within the testing data set. For each threshold, a grid cell at a given time was assigned a value of one, if the precipitation exceeded the threshold, and zero otherwise. Next, we computed the number of true positives (TP; correctly predicted events), false negatives (FN; missed observed events), false positives (FP; predicted events that did not occur), and true negatives (TN; correctly predicted nonevents). Finally, we calculated two skill scores: the Heidke skill score (HSS) and the critical success index (CSI), as defined in Equations 11 and 12 (Zhao et al., 2024). By integrating event-based skill scores with distributional, spatial, and temporal diagnostics, our framework provides

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a uniquely multidimensional evaluation of downscaling performance, going beyond the limited set of metrics typically reported in prior studies. A score of 1 in both metrics indicates perfect similarity, while a score of zero indicates zero similarity.

$$\mathrm{HSS} = \frac{2\,(\mathrm{TP}\cdot\mathrm{TN} - \mathrm{FN}\cdot\mathrm{FP})}{(\mathrm{TP}+\mathrm{FN})\,(\mathrm{FN}+\mathrm{TN}) + (\mathrm{TP}+\mathrm{FP})\,(\mathrm{FP}+\mathrm{TN})} \tag{11}$$

$$\mathrm{CSI} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}} \tag{12}$$

3.5. CMIP6 Application

To demonstrate the applicability of the best-performing model for downscaling CMIP6 simulations, we conduct a two-part analysis that assesses both historical and future performance. First, we evaluate the empirical distributions of downscaled precipitation from the CNRM-ESM2-1 model, generated using the best-performing deep learning model, against those of CanRCM4 during the historical period from 1995 to 2004. This comparison focuses specifically on nonzero hourly precipitation events to ensure a meaningful assessment of the precipitation intensity distributions. Since the deep learning models have been trained on CanRCM4 simulations, the downscaled CMIP6 outputs are expected to exhibit statistical properties that align closely with those of CanRCM4. A strong agreement in the distributional characteristics would indicate that the downscaling model effectively learns and transfers the underlying precipitation patterns, including frequency, intensity, and variability, from the RCM to the GCM outputs.

Second, we extend the analysis to future climate projections by applying the best-performing model to downscale CMIP6 simulations under multiple emission scenarios. This step serves a dual purpose: (a) confirming the expected increasing trends in precipitation associated with a warming climate and (b) evaluating the model's ability to generate "realistic" precipitation fields that match the spatial patterns of the coarse-resolution/raw simulation. The model's ability to reproduce extremes was further evaluated using exceedance probability plots. These plots were created by ranking all nonzero hourly precipitation values from largest to smallest and calculating the proportion of values exceeding each threshold. Specifically, for a given precipitation intensity x, the exceedance probability is defined as the fraction of data points greater than or equal to x. Overall, this evaluation provides a comprehensive assessment of the model's performance in both historical and future contexts, ensuring confidence in its practical applications.

4. Results

4.1. Model Evaluation and Comparison

The performance of the four model configurations was evaluated based on their ability to preserve critical statistical properties, including the probability of zero precipitation (P_0), mean, second L-moment, L-skewness, and L-kurtosis. All models showed reasonable agreement with the original data (Figure 3). While the addition of a nonzero thresholding layer was anticipated to enhance the preservation of P_0, the models with this architectural component (WGAN_NZ and WGAN_NZ_DT) exhibited the highest bias and RMSE in preserving P_0. One possible reason is that the internal thresholding layer may act as a nondifferentiable constraint, disrupting the generator's ability to model fine gradients near the dry/wet boundary and potentially impairing the adversarial training process. In contrast, the simpler WGAN model, with thresholding applied postprediction by setting all values below 0.1 mm to zero, achieved the best performance in preserving P_0, with the lowest bias (−0.006) and RMSE (0.02), followed closely by the UNET model.

For mean precipitation, all the employed models show comparable performance in reproducing the central tendency statistics for the hourly precipitation. The second L-moment results revealed very close bias and RMSE among the WGAN-based models, whereas the UNET model showed the highest bias. Although the more complex WGAN-based models demonstrated better performance in terms of bias and RMSE (metrics that predominantly represent most data points), the dispersion around the diagonal line was slightly greater than the WGAN and UNET models, providing an advantage to these latter models. Increased dispersion around the 1:1 line reflects greater variability in the model's predictive performance, which can be interpreted as higher empirical uncertainty. For higher-order statistical characteristics represented by L-skewness and L-kurtosis, all models performed

10.1029/2025JH000678

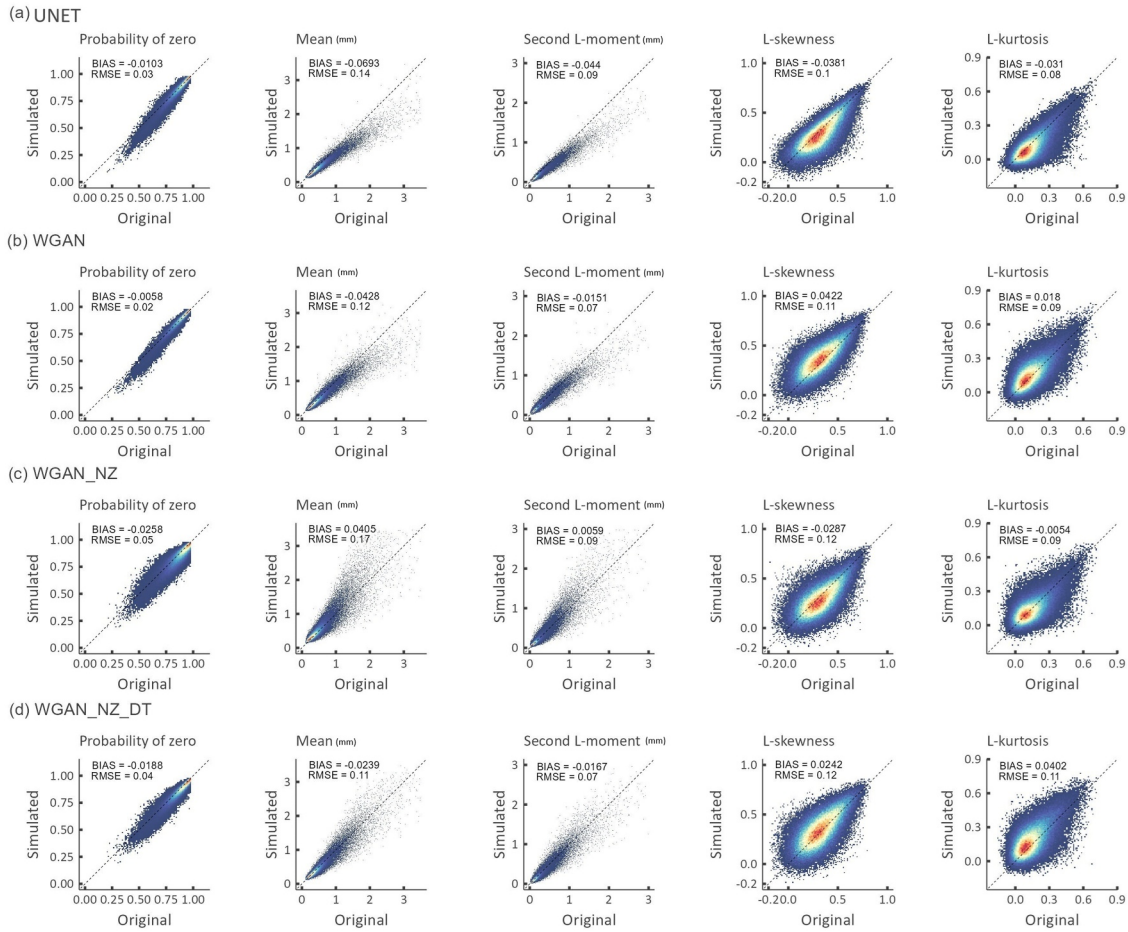

Figure 3. The evaluation between different deep learning models: (a) UNET, (b) WGAN, (c) WGAN_NZ, and (d) WGAN_NZ_DT in preserving the hourly probability of zero and L-moments in the study area. The color scale represents the density of the points; the point density increases as color moves from blue to red.

similarly, with RMSE ranging from 0.10 to 0.12 for L-skewness and from 0.08 to 0.09 for L-kurtosis. These results indicate very comparable bias and RMSE among all employed models. However, the simplest WGAN structure showed the least dispersion among the 1:1 linear compared to the more complex models and slightly better bias and RMSE than the UNET model.

The spatial and temporal correlation performance of the models was assessed to evaluate their ability to reproduce the spatial and temporal structures of the original data (Figure 4). All models captured the general decline in correlation with increasing longitudinal distance (Figures 4a–4c). The WGAN models showed closer alignment with the original simulations, consistently outperforming the UNET model (Figure 4a). The maximum bias observed among the WGAN models was approximately 0.05, with the simplest WGAN model achieving the lowest RMSE. A similar trend was observed for the latitudinal distance lag (Figures 4d–4f). However, the bias and RMSE in this direction were generally smaller than those for the longitudinal direction. This indicates that latitudinal spatial variations are more straightforward for the models to capture, potentially due to the influence of factors such as elevation gradients or simpler precipitation patterns in the latitudinal direction (More interpretation of this finding is in Section 5).

Another metric used to quantify spatial similarity is the similarity index, where higher values indicate greater similarity (refer to Figure S3 in Supporting Information S1). Across all models, the average similarity index between simulated and original fields remained high (0.98), demonstrating overall strong agreement. However, as model complexity increases from UNET to WGAN, WGAN_NZ, and WGAN_NZ_DT, the percentage of fields with a similarity index below 0.8 increased, reaching 0.04%, 0.18%, 0.23%, and 0.30%, respectively. This trend

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000678

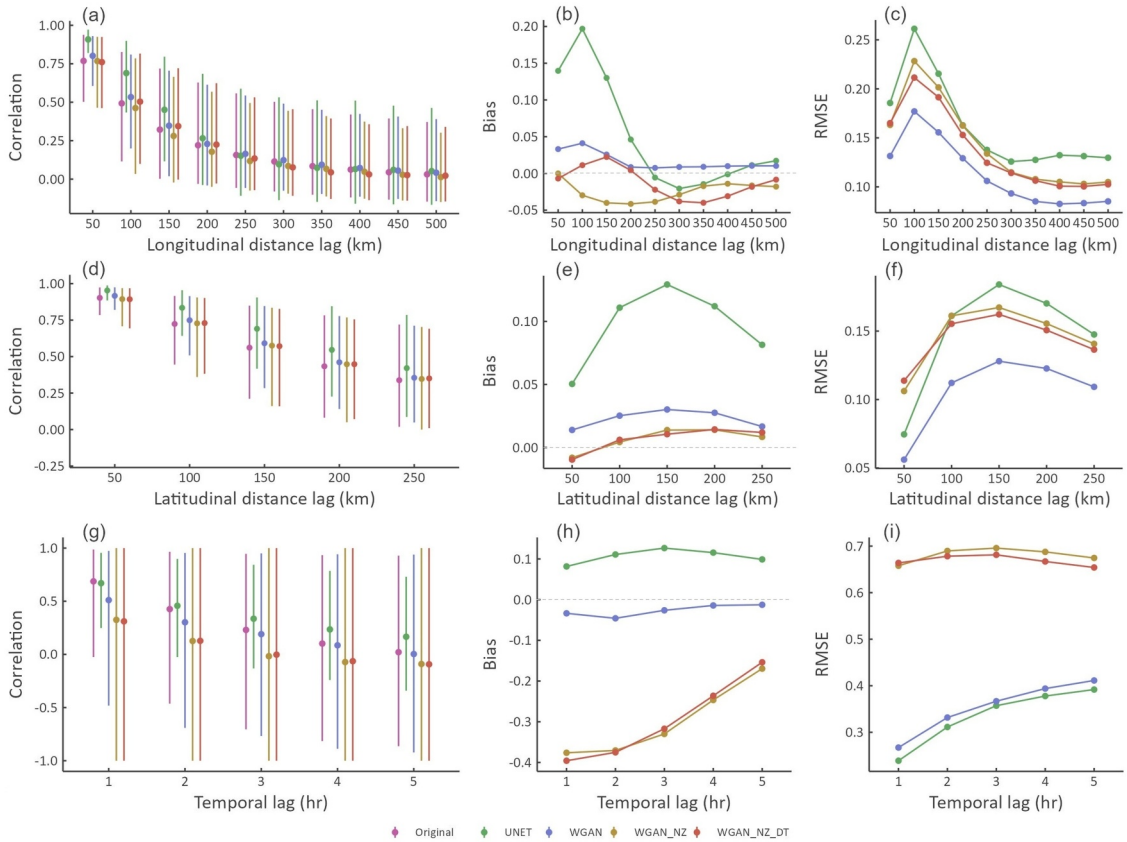

Figure 4. The spatial and temporal correlation of the four proposed models against the original simulations of CanRCM4. (a–c) show the longitudinal spatial correlation results; (d–f) show the latitudinal spatial correlation results; and (g–i) show the autocorrelation structures. The error bars show the empirical 90% confidence interval, while the points show medians.

suggests that while the more complex models generally maintain high similarity, there is a slight increase in the number of outlier fields where the similarity index falls below 0.8.

In terms of temporal correlation, the more complex WGAN_NZ and WGAN_NZ_DT models exhibited higher biases in the negative direction and the largest RMSE values (Figures 4g–4i). This finding suggests that while these models excel in capturing spatial patterns, their complexity may introduce challenges in accurately reproducing temporal relationships. In contrast, the WGAN and UNET models demonstrated better performance, showing closer alignment with the original data's autocorrelation across all lags. Among these, the WGAN model was particularly noteworthy, achieving a bias that was very close to zero (Figure 4h). This observation highlights its ability to balance simplicity and performance, maintaining temporal autocorrelation patterns effectively without the trade-offs seen in more complex models.

To eliminate regional variability and evaluate the temporal dependence of the study region as a whole, several metrics were employed to assess the daily average temporal dependence of the simulated fields compared to the original fields. These metrics include the coherence coefficient, the Kolmogorov-Smirnov (KS) statistic, the autocorrelation function (ACF) bias, and the ACF RMSE (Figure 5). All models demonstrate reasonable coherence, with median values consistently above 0.5. The WGAN model achieves the highest median coherence, followed closely by the UNET model. However, the more complex models, WGAN_NZ and WGAN_NZ_DT, exhibit lower median values. Similarly, for the KS statistic, the UNET and WGAN models show smaller median values than WGAN_NZ and WGAN_NZ_DT, indicating better alignment with the original data distribution. In contrast, the WGAN_NZ and WGAN_NZ_DT models display larger KS statistics with greater variability, reflecting reduced performance in this metric.

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000678

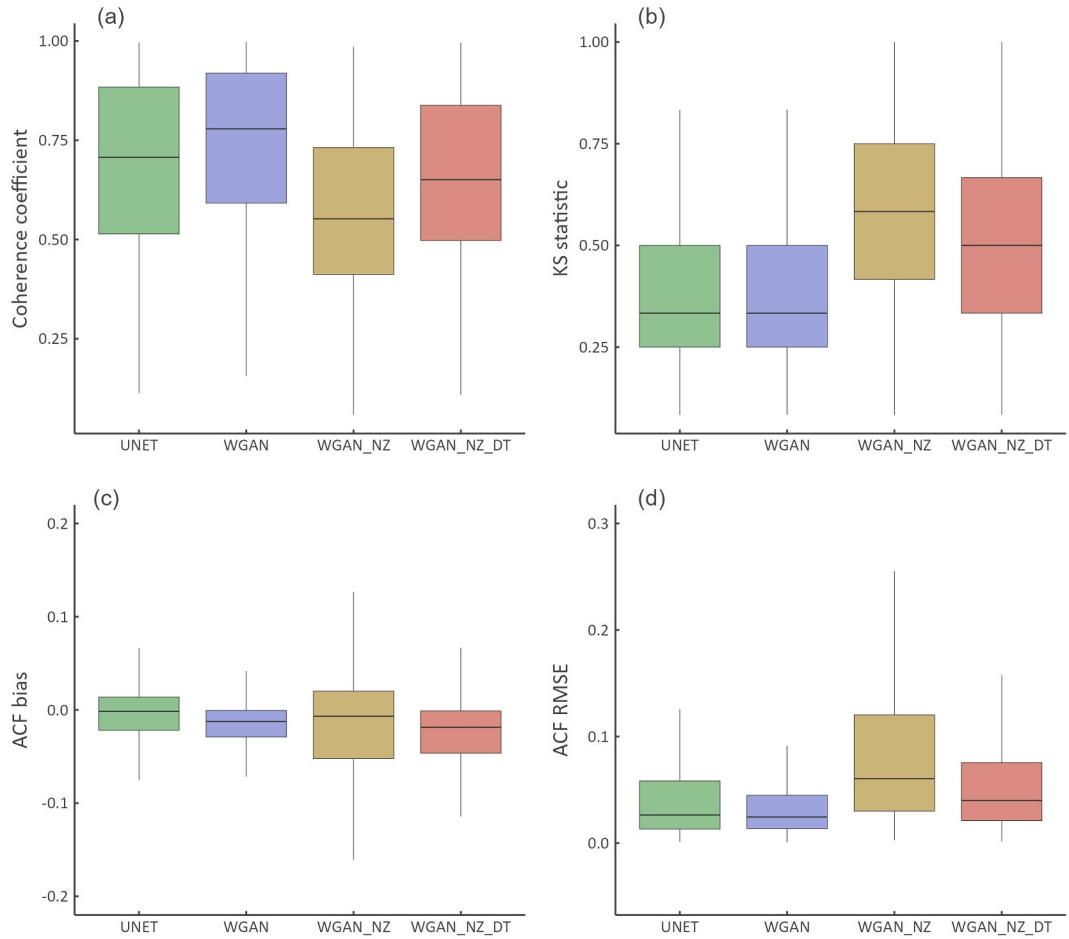

Figure 5. Comparison of the daily (24-hr time series) average temporal dependencies within the study region, derived from the simulated fields and compared to the original fields.

For ACF bias, all models demonstrate smaller biases centered around zero, indicating that none of the models introduce significant bias into temporal dependencies. Among the models, the WGAN model has the narrowest interquartile range and the lowest variability in ACF bias, suggesting more consistent performance in preserving temporal autocorrelation. Conversely, WGAN_NZ shows slightly greater variability, reflecting less reliable reproduction of temporal patterns. Similar results are observed for ACF RMSE, where the UNET and WGAN models outperform the more complex WGAN_NZ and WGAN_NZ_DT models. These findings suggest that increasing model complexity does not necessarily enhance the replication of temporal dependencies and, in some cases, may even reduce performance.

Finally, we evaluated the performance of the three WGAN variants using event-based metrics (HSS and CSI) applied to thresholds corresponding to the median and the 75th percentile of observed hourly precipitation during the first year of the testing period. All WGAN models exhibited comparable performance on average, with HSS (CSI) values ranging from 0.56 to 0.64 (0.43–0.51) for the median threshold and 0.51–0.54 (0.38–0.41) for the 75th percentile. Histograms of these results are presented in Figures S4–S9 in Supporting Information S1. These results demonstrate the ability of all CNN-based deep learning models to accurately capture precipitation events in both space and time. Notably, the simplest WGAN architecture slightly outperformed its more complex counterparts in this event-based evaluation.

To facilitate clearer comparison across the different deep learning models and evaluation metrics, a Table S2 in Supporting Information S1 has been color-coded to enable quick and intuitive interpretation of the results.

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000678

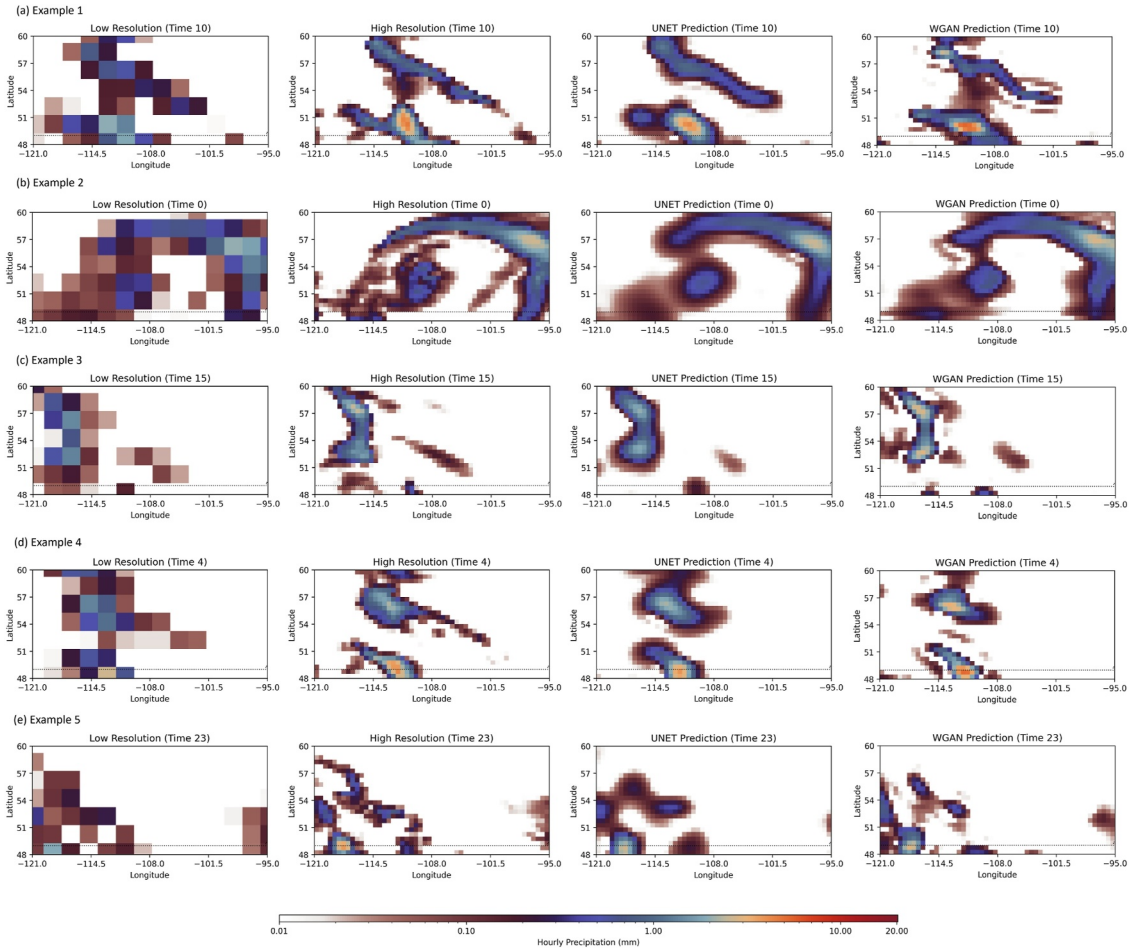

Figure 6. Examples of one-hour precipitation fields generated by the U-Network model (third column) and the Wasserstein generative adversarial networks model (fourth column), compared to the input low-resolution field (first column) and the original high-resolution field (second column).

4.2. WGAN Performance Visualization

To visualize the results of the WGAN model, which demonstrates a balanced and overall stronger performance, we present examples of a one-hour precipitation field generated by the UNET and WGAN models (Figure 6). The UNET model produces a smooth representation of the low-resolution input field, lacking realistic spatial variability and fine-scale features (i.e., heterogeneous and scattered nature of precipitation fields, exhibiting some clustering but without an idealized smoothing effect). In contrast, the WGAN model generates precipitation fields with more realistic spatial patterns that better capture the high-resolution details observed in the original field. Additional examples of generated precipitation fields across different time steps are provided in Figure S10 in Supporting Information S1, further demonstrating the spatial and temporal consistency of the WGAN model. Further, additional examples of 24-hr time series are provided in Figure S11 in Supporting Information S1.

4.3. Downscaling CMIP6 in the Canadian Prairies

As an initial showcase, we applied the UNET and WGAN models to downscale the historical simulation for the period 1995–2004, comparing the results with CanRCM4 simulations for the same time frame. Note that the CNRM-ESM2-1 outputs are offset by 30 min, starting the day at 00:30 UTC, whereas CanRCM4 begins at 00:00 UTC. As a result, discrepancies may be observed in the hourly fields (Figures 7a–7c). However, when comparing the distributions of nonzero hourly precipitation (Figure 7d: exceedance probability plot), the downscaled data align well with the probability distribution of CanRCM4. This alignment in the nonexceedance probability plot provides a more precise comparison of the central portion of the distribution. Looking at extremes, the

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000678

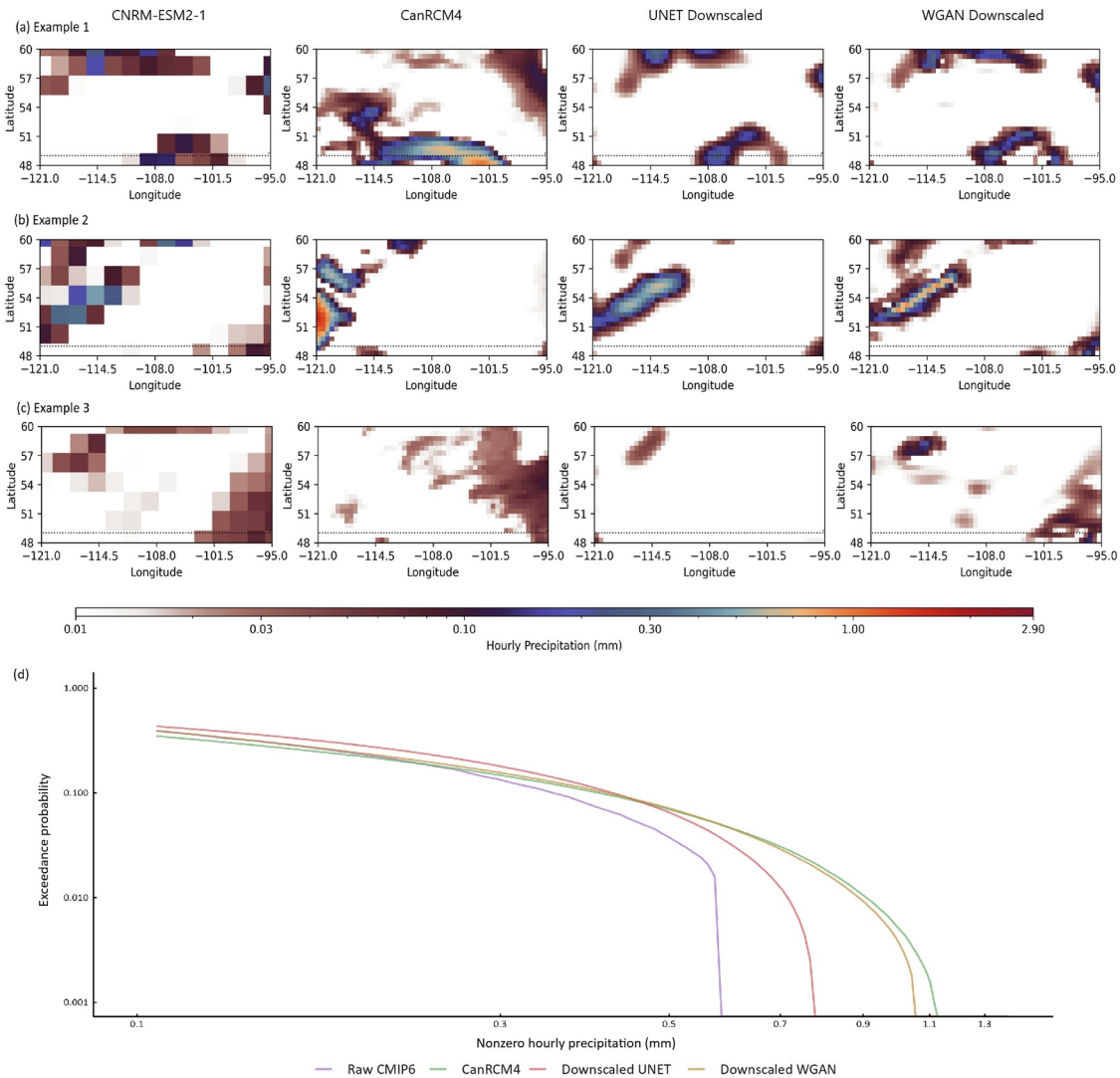

Figure 7. Validating historical CNRM-ESM2-1 simulation (raw Coupled Model Intercomparison Project phase 6 model simulation) with CanRCM4 outputs, showing (a–c) three examples of hourly fields (Note that the CNRM-ESM2-1 outputs are offset by 30 min) and (d) comparing the exceedance probability distributions, showing the median across all fields in the testing data set.

downscaled historical simulation using the WGAN model closely matches the tail of the distribution of the CanRCM4 data. This finding is expected, as the WGAN model is trained on CanRCM4 data, inherently reproducing its statistical properties. In contrast, the UNET model tends to smooth the original field, explaining why the extreme values of the UNET-downscaled data are closer to those of the raw CMIP6 data.

While both models aim to enhance the spatial resolution of the low-resolution inputs, the WGAN model produces more detailed precipitation fields than the UNET model. The WGAN model demonstrates an ability to reconstruct fine-scale spatial details and realistic variability in all four SSP scenarios (Figure S12 in Supporting Information S1). It captures the heterogeneity of precipitation distribution more effectively than the UNET model. In contrast, the UNET model tends to produce overly smoothed precipitation fields, leading to the loss of small-scale features that are critical for accurately representing precipitation dynamics and extremes. The WGAN model reproduces sharper and more localized precipitation patterns, which are characteristic of convective storms in the Canadian prairies. These features are either smoothed out or completely absent in the UNET outputs.

As shown in Figure 7, the WGAN model maintains its ability to generate spatially heterogeneous precipitation distributions across all four SSP scenarios. Notably, precipitation intensity increases under higher SSP scenarios

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000678

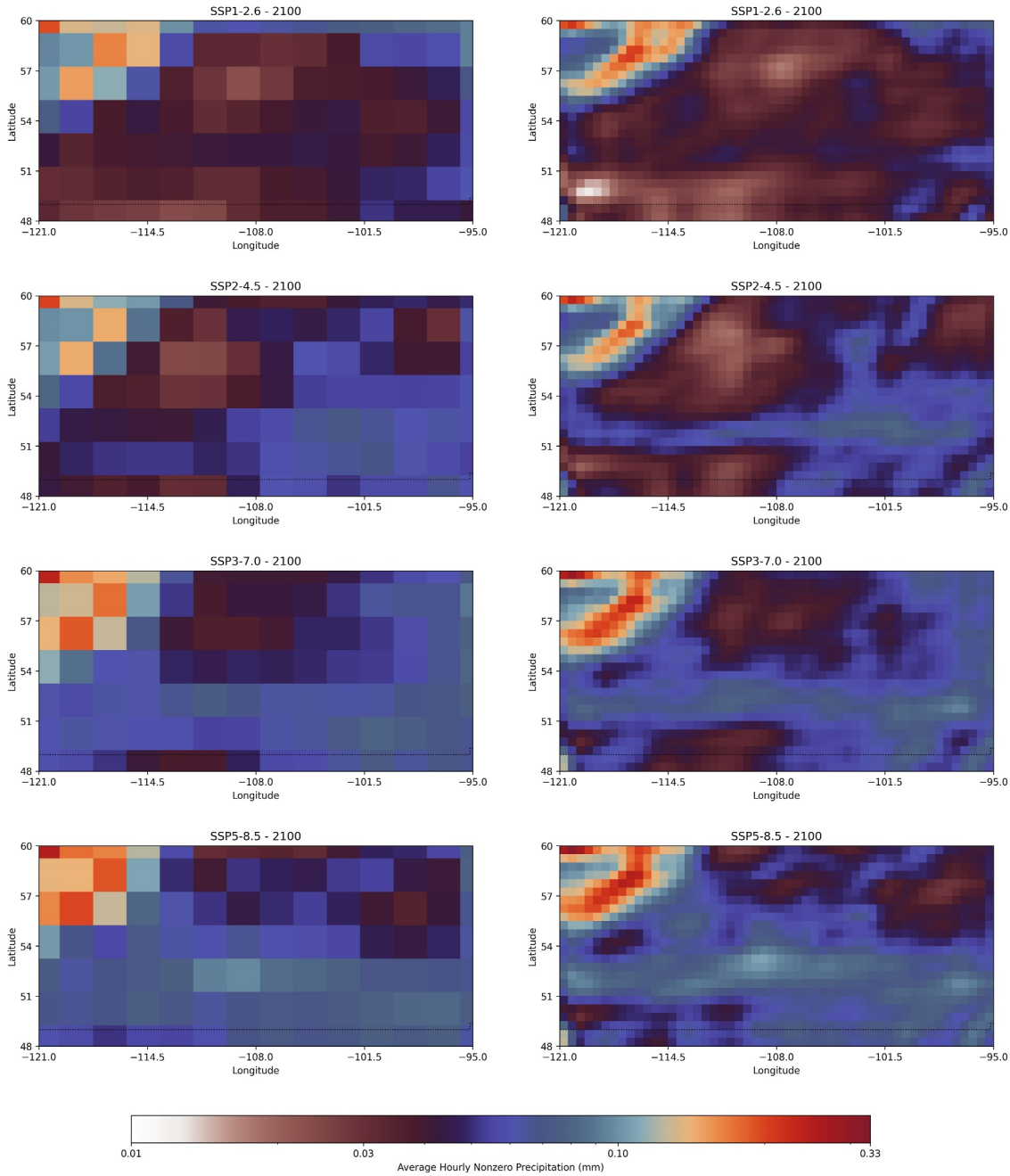

Figure 8. The average hourly nonzero precipitation from the raw CNRM-ESM2-1 model (first column) and the corresponding downscaled field by the Wasserstein generative adversarial networks model (second column), across different scenarios: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 in the last year of the century, 2100. We show the median of 10 generated ensemble members. The distribution of the 10 generated predictions in 2100 is shown in Figure S13 in Supporting Information S1.

by the end of the century (2100), accompanied by more localized patterns and sharper boundaries between different intensity levels of hourly precipitation under climate change (Figure 8). Certain areas, particularly in high-latitude regions, exhibit increased precipitation intensity under future climate scenarios, possibly due to warming-driven moisture availability (Akinremi et al., 1999). The foothills and mountainous areas in the western portion of the domain show stronger precipitation signals in WGAN outputs, reflecting orographic influences that are often lost in lower-resolution fields (Teufel et al., 2017). Additionally, the apparent spatial discontinuity between zero and nonzero precipitation values is anticipated, as hourly precipitation in this Canadian region tends to occur in highly localized patterns (Smith et al., 2022; Vincent & Mekis, 2009). Thus, the WGAN outputs

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

provide a more detailed depiction of precipitation distribution, crucial for applications in hydrology, flood risk assessment, and climate adaptation strategies, especially in regions prone to localized storms like the Canadian prairies.

5. Discussion

The results reveal that all four model configurations perform similarly in reproducing spatial dependence, with a general tendency to capture latitudinal spatial correlation more effectively than longitudinal variability. This discrepancy can be attributed to the temperature gradients and variability in the latitudinal direction, particularly in the Canadian prairies, where hourly precipitation is predominantly driven by convective storms. Convective precipitation events, often triggered by differential heating along latitude lines, exhibit stronger spatial coherence in the latitudinal direction due to the more pronounced influence of temperature gradients (Helmuth et al., 2002; Rind, 1998). Trenberth et al. (2003) emphasized the critical role of temperature gradients in modulating mid-latitude storm tracks and convective activities, particularly in regions influenced by strong seasonal variability like the Canadian prairies. Furthermore, this finding supports the conclusions of Pendergrass and Hartmann (2014), who noted that convective precipitation tends to exhibit higher spatial coherence in regions where temperature gradients dominate precipitation dynamics. By contrast, longitudinal variability is often influenced by more complex factors, including orographic effects, land surface heterogeneity, and wind shear, which introduce greater challenges for downscaling models.

In terms of the temporal dependence, the UNET and WGAN models generally outperform the WGAN_NZ and WGAN_NZ_DT models for reproducing coherence, KS statistic, and autocorrelation metrics. Studies that compared GAN models against the UNET architecture showed that the GAN models produce more detailed and localized outputs (Ravuri et al., 2021). While GANs with additional layers and complex structures can address specific challenges, they often introduce variability and instability, particularly in tasks requiring temporal coherence (Chu et al., 2020). The reduced performance of the WGAN_NZ and WGAN_NZ_DT models in reproducing temporal dependencies may stem from their additional layers' sensitivity to noise and the constraints imposed by the thresholding layer. The thresholding layer, designed to address precipitation intermittency, provides substantial benefits in many applications by restricting the output domain to align with physical reasoning or valid ranges (e.g., Goyal et al., 2023; Hoernle et al., 2021; Konstantinov & Utkin, 2023). However, these layers, in some cases, might restrict the model's flexibility in learning temporal patterns, as suggested by Goodfellow et al. (2016) in their foundational work on GANs. To validate each model's ability to reproduce temporal dependencies, we accounted for training uncertainty by training each configuration 10 times independently. Our findings showed that the WGAN_NZ and WGAN_NZ_DT models exhibited the highest training variability, underscoring their instability compared to WGAN (Figure S14 in Supporting Information S1). This instability is also evident during the training procedure (Figures S15 and S16 in Supporting Information S1). Thus, we found that postprocessing negative values generated by the WGAN model by setting them to zero resulted in better performance than the outputs of the WGAN_NZ model for our specific case study.

Similarly, the ConvLSTM layer in the critic network, while intended to capture temporal evolution, is known to require careful parameter tuning to balance learning efficiency and stability (Shi et al., 2015). The competitive nature of WGAN models, where the generator and the critic are trained adversarially, adds further complexity to finding an optimal configuration for both networks. Including a ConvLSTM layer in the critic may not be strictly necessary, as its primary role is to assess whether the generated fields are realistic. Replacing the ConvLSTM with dense (i.e., fully connected) layers that can process the time dimension explicitly may provide a more numerically stable and efficient alternative. This approach would complement the generator, which already incorporates temporal functionality through the ConvLSTM layer, and may explain why the WGAN without a ConvLSTM layer in the critic performed better. However, further optimization of the network architecture, including the potential integration of LSTM layers, could enhance overall performance. It is crucial to ensure that neither the critic nor the generator dominates the training process, as this imbalance can hinder the performance of the other model and ultimately degrade the system's overall effectiveness. Previous studies have also reported that simpler architectures of GAN-based models may outperform more complex ones (e.g., Huang et al., 2024; Jeong et al., 2022; Purwono et al., 2025).

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

It is worth noting that we initially optimized the WGAN architecture and then introduced the thresholding layer or the ConvLSTM as subtle modifications intended to further improve P_0 and temporal dependence preservation. The lack of improvement suggests that such changes may require a more tailored architectural reoptimization to be effective. GAN-based models are highly sensitive to architectural changes, implying that thresholding mechanisms may need careful integration to yield consistent benefits. Designing WGAN models is not a straightforward process, as it involves considerable case sensitivity and experimentation to achieve the desired outcomes. This inherent complexity underscores the importance of iterative refinement and careful consideration of architectural choices.

The UNET model, as the simplest configuration, is expected to yield results that closely align with the original fields. However, its primary limitation lies in its tendency to act as a smoothing operator, refining the low-resolution input fields into high-resolution outputs but lacking realistic spatial inhomogeneity. This smoothing effect makes it comparable to traditional interpolation techniques, which similarly fail to capture fine-scale variability and realistic spatial patterns. Despite this limitation, the UNET model surpasses interpolation or regression methods in its ability to understand spatial features and temporal dependencies (Liu et al., 2024; Wang et al., 2021). By leveraging deep learning, it can learn and replicate complex patterns within the data that simpler statistical approaches cannot achieve. This ability demonstrates the model's potential as an efficient and straightforward tool for downscaling tasks, though its limitations in representing realistic spatial variability remain.

The effectiveness of the WGAN model in downscaling precipitation fields was further demonstrated by comparing the distribution of historical CMIP6 simulations for the CNRM-ESM2-1 model with CanRCM4. The WGAN model successfully captured the tails of the distribution (i.e., extreme values) with high fidelity to CanRCM4, which was used for training, as also indicated in Figure A1. This result underscores the model's proficiency in preserving spatial and temporal dependencies, as well as key distributional properties, throughout the comprehensive evaluation analyses. Furthermore, when applied to downscale future projections for the year 2100, the WGAN model effectively translated coarse-resolution fields into high-resolution outputs, producing detailed precipitation patterns. The results aligned with expected climate trends, showing an increase in precipitation extremes and a decrease in precipitation averages in the Canadian prairies under a warming climate (Bush & Lemmen, 2019; Mapfumo et al., 2023; McGinn, 2010). This alignment further reinforces the model's reliability for future climate projections.

However, we emphasize that the use of CMIP6 models in this study serves primarily to demonstrate the potential and applicability of WGAN models for downscaling precipitation fields. Further investigation is needed to assess whether models trained on historical data can be reliably applied to future projections and whether nonstationarity is implicitly captured when using GAN-based approaches. Another important consideration is the potential for systematic biases when models are trained on one data set and applied to another, as differences among GCMs and RCMs may influence the robustness of the inference. Quantifying this cross-model uncertainty, such as evaluating how the choice of training RCM affects the generated outputs or how inference on different GCMs behaves without explicit bias correction, remains an open research question. We also note that data-driven models cannot guarantee physical soundness; therefore, incorporating a physical constraint, represented as a mathematical equation, into the loss function can potentially enhance the overall performance of WGAN models. It is important to understand the underlying physical processes governing precipitation and the need for careful architectural choices in model design. However, the results not only highlight the superiority of the WGAN model over simpler approaches like UNET but also demonstrate its potential for improving the reliability of climate projections at regional scales. While this study focuses on the Canadian prairies, we acknowledge that the transferability of the trained WGAN model to other regions with different precipitation regimes may be limited, given its sensitivity to region-specific patterns. Future work could investigate the robustness of the WGAN model under varying climate conditions and its ability to generalize across different climate model outputs for even higher-resolution downscaling.

6. Conclusions

In this study, we compare the performance of four configurations of deep learning models (UNET, WGAN, WGAN_NZ, and WGAN_NZ_DT) for downscaling GCM outputs (~200 km) to the finer spatial resolution of

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

RCMs (~50 km) for hourly precipitation, focusing on the Canadian prairies. The key findings of this study are as follows:

- All models effectively captured spatial dependence due to the use of convolutional networks. However, the simplest WGAN structure demonstrated a slightly better performance.
- Latitudinal spatial correlation was preserved better than longitudinal correlation, which may be attributed to the overall stronger spatial correlation observed in the original fields in the latitudinal direction than the longitudinal direction.
- WGAN models preserved temporal dependence more effectively than the UNET model, which primarily acts as an intelligent smoother. Among the WGAN models, the simplest structure achieved the best results under the given the configuration.
- Introducing a thresholding layer in the generator or a ConvLSTM layer in the critic did not significantly improve performance in this specific case study. However, optimizing the configuration along with ConvLSTM could possibly improve the results.
- Discrepancies between the fields generated by the UNET and WGAN models were further highlighted by downscaling CMIP6 model projections. The UNET model produced overly smoothed fields, whereas the WGAN model generated more localized and detailed fields.
- The WGAN model successfully replicated the distribution of CanRCM4, which was used for training, accurately capturing extreme values.

These findings underscore the potential of deep learning models, particularly WGAN-based architectures, to effectively downscale climate model outputs and generate detailed precipitation fields. Optimizing the architecture further could significantly enhance the reliability of downscaled outputs, improving preparedness, especially in regions like the Canadian prairies, which are prone to intense convective storms. Future studies should focus on refining the spatial resolution of the downscaled fields and exploring diverse patterns of convective storms under various climatic scenarios.

Appendix A: WGAN Stochastic Performance

One of the key advantages of GAN-based models that incorporate noise as an input is their ability to generate ensembles of predictions, enabling probabilistic assessments that are valuable for risk analysis and uncertainty quantification. To evaluate the spread and consistency of simulated outputs relative to the original data, we compare statistical summaries of 100 ensemble members generated by the trained WGAN models with the corresponding values from the original simulations. Figure A1 presents scatterplots of four L-moments and extreme statistics for nonzero hourly precipitation as follows: mean, second L-moment (scale), L-skewness, L-kurtosis, the 99th percentile, and the 99.9th percentile for each day within 1-year period in the testing data set (1987–1988) across the study area.

The strong alignment of the mean ($R^2 = 0.94$) and second L-moment ($R^2 = 0.96$) statistics along the 1:1 line suggests that the model reliably reproduces the central tendency and variability of the original simulations. However, there is more noticeable scatter in the L-skewness ($R^2 = 0.59$) and L-kurtosis ($R^2 = 0.66$) plots, indicating greater uncertainty and less consistency in capturing the shape of the precipitation distribution, particularly in terms of asymmetry and tail heaviness. This deviation implies that while the WGAN ensemble performs well in reproducing the bulk distributional features, its ability to represent extremes is somewhat limited. However, given the inherently uncertain nature of extremes, the fact that most points remain clustered around the 1:1 line still reflects the model's strong potential to capture extreme behavior. Moreover, the WGAN shows an excellent ability to reproduce hourly precipitation extremes in terms of the 99th and 99.9th percentiles within each day across the study area with very high alignment along the 1:1 line ($R^2 = 0.95$ and 0.93, respectively). Overall, the WGAN provides a valuable ensemble-based framework capable of representing a plausible range of extremes, supporting more robust risk assessments in hydrometeorological applications.

10.1029/2025JH000678

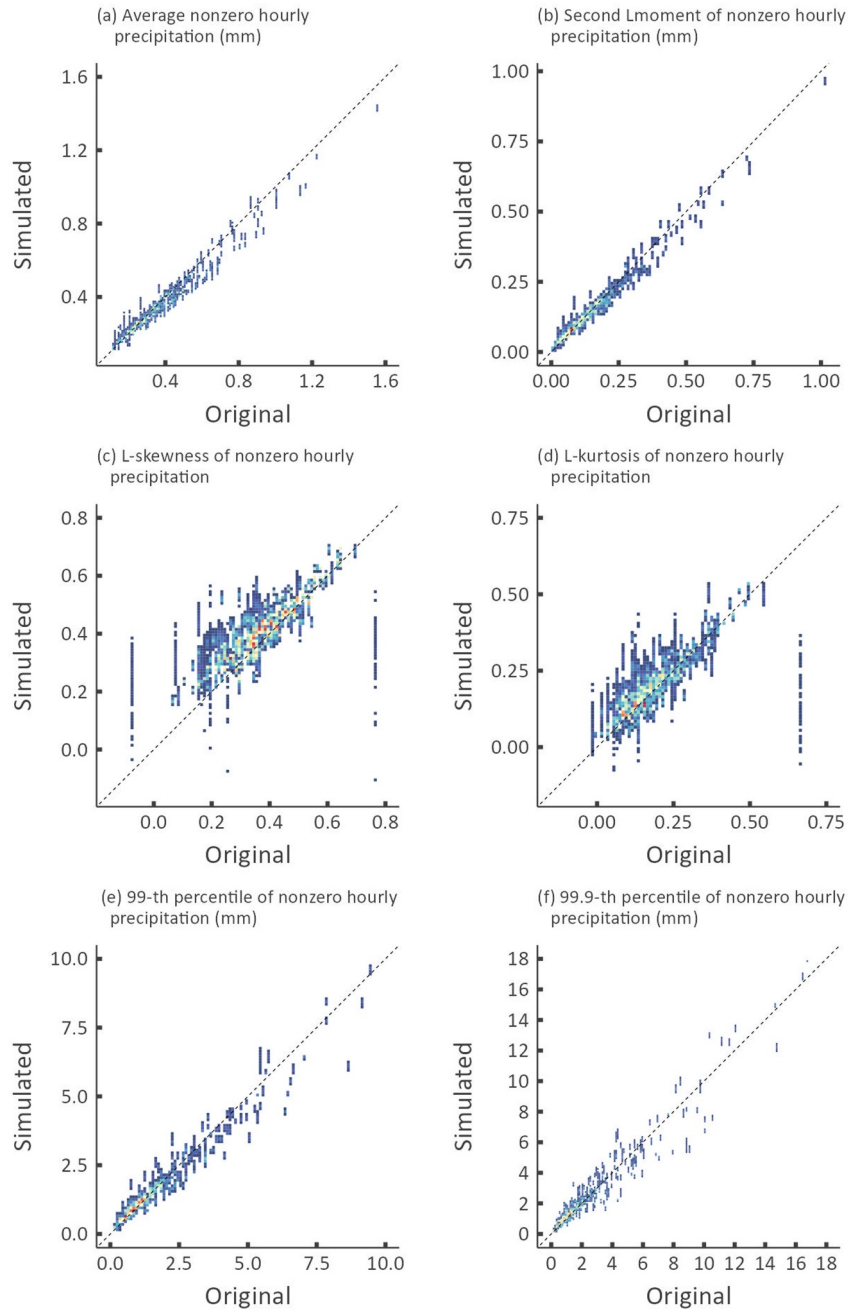

Figure A1. Scatterplots comparing the range of statistics from 100 ensemble predictions generated by the trained Wasserstein generative adversarial networks model against the original simulations for each day and each field, including (a) average, (b) second L-moment, (c) L-skewness, (d) L-kurtosis, (e) the 99th percentile, and (f) the 99.9th percentile of nonzero hourly precipitation over 365 days during the 1987–1988 period. The color scale indicates the density of the points; the red color shows high dense of points, while the blue is less dense of points.

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

CMIP6 simulations provided by ESGF can be found by the following open-source link: https://esgf-node.llnl.gov/search/cmip6/. Users should select the variable as *pr,* which stands for precipitation, select the frequency as *hourly*, select Experiment ID as *historical or ssp126*, *ssp245*, *ssp370*, and *ssp585* for the future scenarios, and select the target CMIP6 model found in this study (i.e., CNRM-ESM2-1 model) and then download the NC files that appear as search outputs (CNRM-CERFACS, 2019).

CanRCM4 simulations are found on the Government of Canada website (2025).

The codes supporting this paper are openly accessible on GitHub at https://doi.org/10.5281/zenodo.17109605 (Abdelmoaty, 2025).

Acknowledgments
H.M.A. is funded by the Mitacs Elevate Program in partnership with BGC Engineering Inc. S.M.P. is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant: RGPIN-2019-06894). A.M. and S.S. gratefully acknowledge the financial support provided by the Environmental Institute, University of Virginia. The authors are sincerely grateful to Professor John Pomeroy for sponsoring H.M.A.'s supercomputer account with the Digital Research Alliance of Canada, through the Global Water Futures Observatories Project at the Global Institute for Water Security. The authors also acknowledge the World Climate Research Programme (WCRP), which coordinated and promoted CMIP6 through its Working Group on Coupled Modeling, the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

References

Abdelmoaty, H. (2025). GAN-based downscaling (version 1.0.0) [Computer software]. *Zenodo*. https://doi.org/10.5281/zenodo.17109605

Akinremi, O. O., McGinn, S. M., & Cutforth, H. W. (1999). Precipitation trends on the Canadian Prairies. *Journal of Climate*, *12*(10), 2996–3003. https://doi.org/10.1175/1520-0442(1999)012<2996:PTOTCP>2.0.CO;2

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), 1–6. https://doi.org/10.1109/icengtechnol.2017.8308186

Annau, N. J., Cannon, A. J., & Monahan, A. H. (2023). Algorithmic hallucinations of near-surface winds: Statistical downscaling with generative adversarial networks to convection-permitting scales. *Artificial Intelligence for the Earth Systems*, *2*(4), e230015. https://doi.org/10.1175/AIES-D-23-0015.1

Baño-Medina, J. (2020). Understanding deep learning decisions in statistical downscaling models. In *Proceedings of the 10th International conference on climate informatics. Presented at the CI2020: 10th international conference on climate informatics, ACM, virtual United Kingdom* (pp. 79–85). https://doi.org/10.1145/3429309.3429321

Bhatia, S., & Dahyot, R. (2019). Using WGAN for improving imbalanced classification performance. In *CEUR Workshop Proceedings* (Vol. 2563, pp. 365–375). CEUR.

Bouvrie, J. (2006). Notes on convolutional neural networks.

Brooks, H. E. (2013). Severe thunderstorms and climate change. *Atmospheric Research*, *123*, 129–138. https://doi.org/10.1016/j.atmosres.2012.04.002

Bush, E., & Lemmen, D. S. (2019). Canada's changing climate report. https://doi.org/10.4095/314614

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Ferro, C. A. T., & Stephenson, D. B. (2012). Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation? *Climate Dynamics*, *41*(5–6), 1475–1495. https://doi.org/10.1007/s00382-012-1568-9

Chaudhuri, C., & Robertson, C. (2020). CliGAN: A structurally sensitive convolutional neural network model for statistical downscaling of precipitation from multi-model ensembles. *Water*, *12*(12), 3353. https://doi.org/10.3390/w12123353

Chen, H., Sun, J., Lin, W., & Xu, H. (2020). Comparison of CMIP6 and CMIP5 models in simulating climate extremes. *Science Bulletin*, *65*(17), 1415–1418. https://doi.org/10.1016/j.scib.2020.05.015

Chu, M., Xie, Y., Mayer, J., Leal-Taixé, L., & Thuerey, N. (2020). Learning temporal coherence via self-supervision for GAN-based video generation. *ACM Transactions on Graphics*, *39*(4). https://doi.org/10.1145/3386569.3392457

CNRM-CERFACS. (2019). CNRM-ESM2-1 model output prepared for CMIP6 [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.1363

Diaconescu, E. P. (2016). Evaluation of precipitation indices over North America from various configurations of regional climate.

Dong, W., Zhao, M., Ming, Y., & Ramaswamy, V. (2021). Representation of tropical mesoscale convective systems in a general circulation model: Climatology and response to global warming. *Journal of Climate*, *34*, 1–40. https://doi.org/10.1175/jcli-d-20-0535.1

Frégier, Y., & Gouray, J.-B. (2021). Mind2Mind: Transfer learning for GANs. *Lecture Notes in Computer Science*, 851–859. https://doi.org/10.1007/978-3-030-80209-7_91

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Government of Canada. (2025). CanRCM4 climate model simulations [Dataset]. *Climate Modelling and Analysis*. Retrieved from https://climate-modelling.canada.ca/climatemodeldata/canrcm/CanRCM4/

Goyal, K., Dumancic, S., & Blockeel, H. (2023). DeepSaDe: Learning neural networks that guarantee domain constraint satisfaction. https://doi.org/10.48550/arXiv.2303.01141

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein GANs. https://doi.org/10.48550/arXiv.1704.00028

Ham, H., Jun, T. J., & Kim, D. (2020). Unbalanced GANs: Pre-training the generator of generative adversarial network using variational autoencoder.

Hanesiak, J., Stewart, R., Painchaud-Niemi, D., Milrad, S., Liu, G., Vieira, M., et al. (2022). The severe multi-day October 2019 snow storm over Southern Manitoba, Canada. *Atmosphere-Ocean*, *60*(2), 65–87. https://doi.org/10.1080/07055900.2022.2060794

Helmuth, B., Harley, C. D. G., Halpin, P. M., O'Donnell, M., Hofmann, G. E., & Blanchette, C. A. (2002). Climate change and latitudinal patterns of intertidal thermal stress. *Science*, *298*(5595), 1015–1017. https://doi.org/10.1126/science.1076814

Hess, P., Lange, S., Schötz, C., & Boers, N. (2023). Deep learning forbias-correcting CMIP6-class Earthsystem models. *Earth's Future*, *11*, e2023EF004002. https://doi.org/10.1029/2023EF004002

Hoernle, N., Karampatsis, R. M., Belle, V., & Gal, K. (2021). MultiplexNet: Towards fully satisfied logical constraints in neural networks. https://doi.org/10.48550/arXiv.2111.01564

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Huang, Y., Gokaslan, A., Kuleshov, V., & Tompkin, J. (2024). The GAN is dead; long live the GAN! A modern baseline GAN.
Hwang, Y., Ma, X., Agyeman, R. Y., & Li, Y. (2023). Climatological features of future MCSs over the Canadian Prairies using convection-permitting climate models. *Atmospheric Research*, *293*, 106922. https://doi.org/10.1016/j.atmosres.2023.106922
Iotti, M., Davini, P., von Hardenberg, J., & Zappa, G. (2025). RainScaleGAN: A conditional generative adversarial network for rainfall downscaling. *Artificial Intelligence for the Earth Systems*, *4*(3). https://doi.org/10.1175/aies-d-24-0074.1
Jeong, J. J., Tariq, A., Adejumo, T., Trivedi, H., Gichoya, J. W., & Banerjee, I. (2022). Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation. *Journal of Digital Imaging*, *35*(2), 137–152. https://doi.org/10.1007/s10278-021-00556-w
Kaparakis, C., & Mehrkanoon, S. (2023). WF-UNet: Weather data fusion using 3D-UNet for precipitation nowcasting. *Procedia Computer Science*, *222*, 223–232. https://doi.org/10.1016/j.procs.2023.08.160
Konstantinov, A. V., & Utkin, L. V. (2023). A new computationally simple approach for implementing neural networks with output hard constraints. *Doklady Mathematics*, *108*(S2), S233–S241. https://doi.org/10.1134/s1064562423701077
Lauer, M., Rinke, A., Gorodetskaya, I., Sprenger, M., Mech, M., & Crewell, S. (2023). Influence of atmospheric rivers and associated weather systems on precipitation in the Arctic. *Atmospheric Chemistry and Physics*, *23*(15), 8705–8726. https://doi.org/10.5194/acp-23-8705-2023
Lei, M., Li, J., Zhang, L., Deng, C., Li, Y., & He, J. (2022). Inconsistent frequency trends between hourly and daily precipitation during warm season in Mainland of China. *Geophysical Research Letters*, *49*(19), e2022GL100277. https://doi.org/10.1029/2022GL100277
Leinonen, J., Nerini, D., & Berne, A. (2021). Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(9), 7211–7223. https://doi.org/10.1109/TGRS.2020.3032790
Li, G., & Cao, G. (2024). Generative adversarial models for extreme geospatial downscaling. https://doi.org/10.48550/arXiv.2402.14049
Li, J., Li, L., Zhang, T., Xing, H., Shi, Y., Li, Z., et al. (2024). Flood forecasting based on radar precipitation nowcasting using U-net and its improved models. *Journal of Hydrology*, *632*, 130871. https://doi.org/10.1016/j.jhydrol.2024.130871
Li, L., & Li, Z. (2023). Potential intensification of hourly precipitation extremes in Western Canada: A comprehensive understanding of precipitation-temperature scaling. *Atmospheric Research*, *295*, 106979. https://doi.org/10.1016/j.atmosres.2023.106979
Liu, J., Shi, C., Ge, L., Tie, R., Chen, X., Zhou, T., et al. (2024). Enhanced wind field spatial downscaling method using UNET architecture and dual cross-attention mechanism. *Remote Sensing*, *16*(11), 1867. https://doi.org/10.3390/rs16111867
Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26. https://doi.org/10.1016/j.neucom.2016.12.038
Luo, N., Guo, Y., Chou, J., & Gao, Z. (2021). Added value of CMIP6 models over CMIP5 models in simulating the climatological precipitation extremes in China. *International Journal of Climatology*, *42*(2), 1148–1164. https://doi.org/10.1002/joc.7294
Mapfumo, E., Chanasyk, D. S., Puurveen, D., Elton, S., & Acharya, S. (2023). Historic climate change trends and impacts on crop yields in key agricultural areas of the Prairie Provinces in Canada: A literature review. *Canadian Journal of Plant Science*, *103*(3), 243–258. https://doi.org/10.1139/cjps-2022-0215
McGinn, S. M. (2010). Weather and climate patterns in Canada's prairie grasslands. In J. Shorthouse (Ed.), *Arthropods of Canadian grasslands: Ecology and interactions in grassland habitats* (Vol. 1, pp. 105–119). Biological Survey of Canada. https://doi.org/10.3752/9780968932148.ch5
Milrad, S. M., Gyakum, J. R., Atallah, E. H., & Smith, J. F. (2011). A diagnostic examination of the Eastern Ontario and Western Quebec wintertime convection event of 28 January 2010. *Weather and Forecasting*, *26*(3), 301–318. https://doi.org/10.1175/2010waf2222432.1
Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. https://doi.org/10.48550/arXiv.1411.1784
O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016
O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.
Pendergrass, A. G., & Hartmann, D. L. (2014). Changes in the distribution of rain frequency and intensity in response to global warming. *Journal of Climate*, *27*(22), 8372–8383. https://doi.org/10.1175/JCLI-D-14-00183.1
Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. *Nature Climate Change*, *7*(1), 48–52. https://doi.org/10.1038/nclimate3168
Purwono, P., Wulandari, A. N. E., Ma'arif, A., & Salah, W. A. (2025). Understanding generative adversarial networks (GANs): A review. *Control Systems and Optimizations Letters*, *3*(1), 36–45. https://doi.org/10.59247/csol.v3i1.170
Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. https://doi.org/10.48550/arXiv.1511.06434
Rampal, N., Gibson, P. B., Sherwood, S., Abramowitz, G., & Hobeichi, S. (2024). A robust generative adversarial network approach for climate downscaling and weather generation. https://doi.org/10.22541/essoar.171352077.78968815/v1
Rampal, N., Gibson, P. B., Sherwood, S., Abramowitz, G., & Hobeichi, S. (2025). A reliable generative adversarial network approach for climate downscaling and weather generation. *Journal of Advances in Modeling Earth Systems*, *17*(1), e2024MS004668. https://doi.org/10.1029/2024MS004668
Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, *597*(7878), 672–677. https://doi.org/10.1038/s41586-021-03854-z
Rebora, N., Ferraris, L., Von Hardenberg, J., & Provenzale, A. (2006). RainFARM: Rainfall downscaling by a filtered autoregressive model. *Journal of Hydrometeorology*, *7*(4), 724–738. https://doi.org/10.1175/JHM517.1
Rind, D. (1998). Latitudinal temperature gradients and climate change. *Journal of Geophysical Research*, *103*(D6), 5943–5971. https://doi.org/10.1029/97JD03649
Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Girolamo, S. D., et al. (2020). Kilometer-scale climate models.
Scinocca, J. F., Kharin, V. V., Jiao, Y., Qian, M. W., Lazare, M., Solheim, L., et al. (2016). Coordinated global and regional climate modeling. *Journal of Climate*, *29*(1), 17–35. https://doi.org/10.1175/jcli-d-15-0161.1
Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. https://doi.org/10.48550/arXiv.1506.04214
Shook, K., Papalexiou, S., & Pomeroy, J. W. (2021). Quantifying the effects of Prairie depressional storage complexes on drainage basin connectivity. *Journal of Hydrology*, *593*, 125846. https://doi.org/10.1016/j.jhydrol.2020.125846
Smith, C. D., Mekis, E., Hartwell, M., & Ross, A. (2022). The hourly wind-bias-adjusted precipitation data set from the Environment and Climate Change Canada automated surface observation network (2001–2019). *Earth System Science Data*, *14*(12), 5253–5265. https://doi.org/10.5194/essd-14-5253-2022
Song, Y. H., Nashwan, M. S., Chung, E.-S., & Shahid, S. (2021). Advances in CMIP6 INM-CM5 over CMIP5 INM-CM4 for precipitation simulation in South Korea. *Atmospheric Research*, *247*, 105261. https://doi.org/10.1016/j.atmosres.2020.105261

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sun, Y., Deng, K., Ren, K., Liu, J., Deng, C., & Jin, Y. (2024). Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *208*, 14–38. https://doi.org/10.1016/j.isprsjprs.2023.12.011

Teufel, B., Diro, G. T., Whan, K., Milrad, S. M., Jeong, D. I., Ganji, A., et al. (2017). Investigation of the 2013 Alberta flood from weather and climate perspectives. *Climate Dynamics*, *48*(9–10), 2881–2899. https://doi.org/10.1007/s00382-016-3239-8

Thomassen, E. D., Thorndahl, S. L., Andersen, C. B., Gregersen, I. B., Arnbjerg-Nielsen, K., & Sørup, H. J. D. (2022). Comparing spatial metrics of extreme precipitation between data from rain gauges, weather radar and high-resolution climate model re-analyses. *Journal of Hydrology*, *610*, 127915. https://doi.org/10.1016/j.jhydrol.2022.127915

Tirel, L., Ali, A. M., & Hashim, H. A. (2024). Novel hybrid integrated Pix2Pix and WGAN model with gradient penalty for binary images denoising. *Systems and Soft Computing*, *6*, 200122. https://doi.org/10.1016/j.sasc.2024.200122

Trebing, K., Stańczyk, T., & Mehrkanoon, S. (2021). SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognition Letters*, *145*, 178–186. https://doi.org/10.1016/j.patrec.2021.01.036

Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of precipitation. *Bulletin of the American Meteorological Society*, *84*(9), 1205–1218. https://doi.org/10.1175/BAMS-84-9-1205

Vincent, L. A., & Mekis, E. (2009). Discontinuities due to joining precipitation station observations in Canada. *Journal of Applied Meteorology and Climatology*, *48*(1), 156–166. https://doi.org/10.1175/2008JAMC2031.1

Wang, F., Tian, D., Lowe, L., Kalin, L., & Lehrter, J. (2021). Deep learning for daily precipitation and temperature downscaling. *Water Resources Research*, *57*(4), e2020WR029308. https://doi.org/10.1029/2020WR029308

Wiatrak, M., Albrecht, S. V., & Nystrom, A. (2020). Stabilizing generative adversarial networks: A Survey. https://doi.org/10.48550/arXiv.1910.00927

Zhao, Z., Dong, X., Wang, Y., & Hu, C. (2024). Advancing realistic precipitation nowcasting with a spatiotemporal transformer-based denoising diffusion model. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–15. https://doi.org/10.1109/TGRS.2024.3355755

Zhou, W., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

29935210, 2025, 4, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000678 by Technische Universität Hamburg Universitatsbibliothek, Wiley Online Library on [12/11/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License